Spatial DCT-Based Channel Estimation in Multi-Antenna Multi-Cell Interference Channels

Maha Alodeh, *Student Member, IEEE*, Symeon Chatzinotas, *Senior Member, IEEE*,
Björn Ottersten, *Fellow Member, IEEE*

Abstract—This work addresses channel estimation in multiple antenna multicell interference-limited networks. Channel state information (CSI) acquisition is vital for interference mitigation. Wireless networks often suffer from multicell interference, which can be mitigated by deploying beamforming to spatially direct the transmissions. The accuracy of the estimated CSI plays an important role in designing accurate beamformers that can control the amount of interference created from simultaneous spatial transmissions to mobile users. Therefore, a new technique based on the structure of the spatial covariance matrix and the discrete cosine transform (DCT) is proposed to enhance channel estimation in the presence of interference. Bayesian estimation and Least Squares estimation frameworks are introduced by utilizing the DCT to separate the overlapping spatial paths that create the interference. The spatial domain is thus exploited to mitigate the contamination which is able to discriminate across interfering users. Gains over conventional channel estimation techniques are presented in our simulations which are also valid for a small number of antennas.

Index Terms—Channel estimation, training sequence contamination, discrete cosine transform, second order statistics.

I. INTRODUCTION

Interference is the most critical factor for designing and scaling wireless networks, as it leads to the spectrum scarcity-congestion problem. Moreover, the design paradigm for cellular networks has been shifted from partial frequency reuse to full frequency reuse enhancing the spectrum utilization, and thus making the problem of interference more acute. Therefore, the design of future networks will require collaborating base stations to jointly serve their users or to smartly mitigate the interference. This can be enabled by exchanging data information and channel state information (CSI). These designs have received much attention in the literature, but their main drawback is the requirement of backhaul exchange for users' data, which requires major upgrades of current infrastructure especially when CSI changes rapidly [1]- [3]. In an effort to tackle the interference issue, without data sharing over the backhaul network, various coordination techniques has been proposed to handle the limited data exchange scenarios; for

Maha Alodeh, Symeon Chantzinotas and Björn Ottersten are with Interdisciplinary Centre for Security Reliability and Trust (SnT) at the University of Luxembourg, Luxembourg. E-mails:{ maha.alodeh@uni.lu, symeon.chatzinotas @uni.lu and bjorn.ottersten@uni.lu}. B. Ottersten is also with Royal Institute of Technology (KTH), Stockholm, Sweden. E-mail{bjorn.ottersten@ee.kth.se}.

example the authors in [1] exploit the availability of CSI at base stations to design precoding techniques that minimize the interference created by BS transmissions by maximizing the signal to leakage noise ratio (i.e. virtual signal to interference noise ratio), while the work in [3] investigates zero forcing beamforming in a coordinated multicell environment. Deploying these techniques requires accurate channel state information to design the suitable transmit and receive beamforming.

The CSI acquisition techniques can be categorized into feedback and reciprocity techniques. In the feedback systems, a training sequence is broadcasted by the BS which is measured by users and a limited feedback link is considered from the users to the base station. In [4]- [7] such a mode is considered. In [4], the authors have proved that in order to achieve full multiplexing gain in the MIMO downlink channel in the high signal to noise ratio (SNR) regime, the required feedback rate per user grows linearly with the SNR (in dB). The main result in [5] is that the extent of CSI feedback can be reduced by exploiting multi-user diversity. While in [6], it is shown that non-random vector quantizers can significantly increase the MIMO downlink throughput. In [5], the authors design a joint CSI quantization, beamforming and scheduling algorithm to attain optimal throughput scaling. Authors in [7] present an investigation of how many feedback bits per user are necessary to maintain the optimal multiplexing gain in a K-cell MIMO interference channel. In the second mode, the distinguishing feature of such systems is the concept of reciprocity, where the uplink channel is utilized as an estimate of the downlink channel reducing the feedback requirements. This is one of the main advantages of a TDD architecture in low mobility scenarios [8], as it eliminates the need for feedback, and joint uplink training combined with the reciprocity of the wireless medium are sufficient to estimate the desired CSI. TDD has the advantage over direct feedback since the users' terminal do not need to estimate their own channel. In TDD systems, the base station estimates the channel state information (CSI) based on uplink training sequences over the same frequency band, and then uses it to generate the beamforming vector in the downlink transmissions [9]- [11].

CSI is typically acquired by exploiting finite-length training sequences in the presence of inter-cell interference. Therefore, in a synchronous setting, the training sequences from neighboring cells would contaminate each other. While in an asynchronous setting, the training sequences are contaminated by the downlink data transmissions. Recently, the problem of non-orthogonality of training sequences has been thoroughly

investigated [14]- [20]. It is pointed out in [14] that training sequence contamination presents a huge challenge for performance and a robust precoding technique is proposed to handle this kind of interference. Specifically, it is shown that training sequence contamination effects [16]- [17], [20]- [21] (i.e., the reuse of non-orthogonal training sequences across interfering cells) causes the interference rejection performance to quickly saturate with the number of antennas, thereby undermining the value of MIMO systems in cellular networks [16].

The problem of training sequence contamination can occur in two different scenarios:

- Inter-cell training sequence contamination: presents the most common scenario in which the users in the same cell utilize orthogonal training sequences while this set is reused in other cells.

- Intra-cell training sequence contamination: most research ignores this kind of contamination as they assume no training sequence reuse within the same cell and the training sequences are orthogonal. However, it is worth investigating the scenario in which the number of simultaneously tackled transmissions in single cell is greater than number of training sequences [23].

Although we mention two types of contamination, handling the contamination in both scenarios is the same. We focus on the first type of contamination in this work. Various contamination mitigation techniques have been investigated in the literature; a novel asynchronous training sequence transmission is proposed in [15], where different time slots are allocated to the users who utilize the same training sequence in different cells. The channel estimates are decontaminated from the downlink interference by adding additional antennas to estimate and cancel the interference. On the other hand, the effect of pilot contamination is thoroughly studied in the literature and been characterized in [16]- [17] using large system analysis for different channels. Training sequence allocation techniques are proposed in [21]- [22], to find the optimal set of users that simultaneously utilizes the same training sequence. The allocation schemes vary according to the considered scenario. The result in [21] depends on the channel power and strongest interferer while the work in [22] relies on second order statistics to pick the users who have the lowest overlap in their subspaces.

In this paper, we tackle the problem of training sequence contamination in correlated multiple user single input multiple output (MU-SIMO) in the uplink of TDD scenarios. Although the multiple antennas at users' terminal provide additional degrees of freedom, they present a source of additional complexity from practical perspective which is preferable to be tackled at the BS. Moreover, the assumption of exploiting single antenna at the users' terminal can be motivated by the fact that multiplexing multiple streams per users is less beneficial than sending single stream exploiting receive combining techniques [13]. Based on this observation, we assume a uniform linear array (ULA) at the base station and exploit the embedded information in second order statistics about users' favorable and unfavorable directions.

From a different perspective, orthogonal transformations are exploited in the literature to virtually represent the information in a different domain which simplifies the analysis. In [24], a channel modelling problem for a single user multiple input multiple output (SU-MIMO) is investigated using orthogonal transformations to provide a geometric interpretation of the scattering environment. This virtual transformation reveals two important aspects: the number of parallel channels and the level of diversity and clarify their impact on capacity calculations. In this paper, we exploit the orthogonal transformations' capability of virtually converting the information into a different domain and based on this we propose several signal processing algorithms. One of the orthogonal transformations that has appealing characteristics is the discrete cosine transform (DCT). DCT is utilized to compress the dimensions used for channel estimation, thus facilitating interference mitigation at the estimation process. The contributions of this paper can be summarized as

- Improving the contamination mitigation performance of the traditional Bayesian estimation (BE) at the estimation step. The authors in [22] exploit BE to mitigate the contamination and they analyzed its performance using large system analysis. In this work, we combine BE with DCT in different algorithms to decontaminate the training sequences at the estimation process. Moreover, a modification is employed on traditional BE to further improve the contamination rejection from the target estimate. Numerically, the proposed algorithms are shown to outperform the estimation algorithm proposed in [22].

- Traditionally, least squares estimation (LS) is not capable of discriminating the interference. We apply LS in the DCT domain to mitigate the contamination at the estimation process by exploiting DCT characteristics.

- Training sequence allocation techniques are proposed to find the optimal set of users that utilize the same training sequences. In an effort to utilize the multiuser diversity concept as in [21]- [22], we propose joint training sequence allocation and DCT compression to combine the benefit of both schemes. These allocation techniques further suit the nature of enhanced LS and BE thereby outperforming the traditional techniques.

A. Notation

The adopted notations in the paper are as follows: we use uppercase and lowercase boldface to denote matrices and vectors. Specifically, $\mathbf{I}_K$ denotes the $K \times K$ identity matrix. Let $\mathbf{X}^T$, $\mathbf{X}^*$ and $\mathbf{X}^H$ denote the transpose, conjugate, and conjugate transpose of a matrix $\mathbf{X}$ respectively. $\mathbb{E}$ refers to the expectation, $\| \cdot \|$ denotes the Frobenius norm, and $\| \cdot \|_0$ denotes the zero norm. The Kronecker product of two matrices $\mathbf{X}$ and $\mathbf{Y}$ is denoted by $\mathbf{X} \otimes \mathbf{Y}$. The notation used for $\triangleq$ is used for definitions. $\mathbf{1}^{K \times K}$, $\mathbf{0}^{K \times K}$ are the matrices of all ones and zeros with size $K \times K$ respectively, $\odot$ denotes Hadamard product and $[\mathbf{A}(k, n)]$ is $(k, n)^{th}$ element of matrix $\mathbf{A}$. Let $tr(\mathbf{X})$, $vec(\mathbf{X})$ denote the trace operation, and the column vector obtained by stacking the columns of $\mathbf{X}$. $\mathcal{CN}(a, \mathbf{R})$ is used to denote circularly symmetric complex Gaussian random

vectors, which has the mean a and the covariance matrix $\mathbf{R}$. Finally, $\cup$ is the union of sets and $f(x)$ denotes a function of x.

II. SYSTEM MODEL

Our model consists of a network of C time-synchronized cells with full spectrum reuse, each one of the cells serves L users. Estimation of flat block fading, narrow band channels in the uplink is considered, and all the base stations are equipped with an M-element uniform linear array (ULA) of antennas. We assume that the training sequences, of length τ symbols, used by single-antenna users in the same cell are mutually orthogonal. However, training sequences are reused in a multicell environment from cell to cell. The training sequences used for estimating the user channels are denoted by $\mathbf{s}_i \triangleq [s_{i1} \ \dots \ s_{i\tau}]^T \in \mathbb{C}^{\tau \times 1}$. The training sequence symbols are normalized such that $\{|s_{ij}|^2 = \frac{P}{\tau}, \forall j \in \tau\}$, where P is the total training sequence power. Channel vectors are assumed to be $\mathbb{C}^{M \times 1}$ Rayleigh fading with correlation due to the finite multipath angle spread seen from the base station side, the lth user's channel towards c^{th} BS is given by $\mathbf{h}_{lc} \sim \mathcal{CN}(0, \alpha_{lc}\mathbf{R}_{lc}) \in \mathbb{C}^{M \times 1}$, where α_{lc} is the attenuation from the lth user to c^{th} BS. We denote the channel covariance matrix $\mathbf{R}_{lc} \in \mathbb{C}^{M \times M}$ as $\mathbf{R}_{lc} = \mathbb{E}[\mathbf{h}_{lc}\mathbf{h}_{lc}^H]$.

Considering the transmission of $\mathbf{s}_i$ sequence, the $M \times \tau$ signal baseband symbols sampled at the cth target base station is

$$\mathbf{Y}_c = \sum_i \sum_{\forall l \in \mathcal{K}_i} \mathbf{h}_{lc}\mathbf{s}_i^T + \mathbf{N}_c. \tag{1}$$

where $\mathcal{K}_i$ is the set of users who use the training sequence $\mathbf{s}_i$. $\mathbf{N}_c \in \mathbb{C}^{M \times \tau}$ is the spatially and temporally white complex additive Gaussian noise (AWGN) with element-wise variance σ^2. We define a training matrix $\mathbf{S}_i = \mathbf{s}_i \otimes \mathbf{I}_M$, such that $\mathbf{S}_i^H \mathbf{S}_i = \tau \mathbf{I}_M$. Then, the received training signal at the target base station can be expressed as

$$\mathbf{y}_c = vec(\mathbf{Y}_c) = \sum_i \mathbf{S}_i \sum_{\forall l \in \mathcal{K}_i}^{C} \mathbf{h}_{lc} + \mathbf{n}_c \tag{2}$$

where $\mathbf{n}_c \in \mathbb{C}^{M\tau \times 1} = vec(\mathbf{N})$ is the sampled noise at the cth BS. Since the pilots are orthogonal $\mathbf{s}_i^H \mathbf{s}_j = 0$, this makes $\mathbf{S}_i^H \mathbf{S}_j = \mathbf{0}^{M \times M}$. Due to orthogonality between different sequences, the sampled signal resulted from the transmissions of the ith training sequence at c^{th} BS can be isolated from other training sequences and can be expressed as

$$\mathbf{y}_c = \mathbf{S}_i \sum_{\forall l \in \mathcal{K}_i} \mathbf{h}_{lc} + \mathbf{n}_c \tag{3}$$

For the sake of simplicity and without lack of generalization, we drop the training sequence index and assume that a single training sequence is used over the network, which makes the baseband signal sampled at c^{th} base station as

$$\mathbf{y}_c = \mathbf{S} \sum_{l=1}^{C} \mathbf{h}_{lc} + \mathbf{n}_c \tag{4}$$

where $l = 1, \dots, C$ denotes the users transmitting the training sequence $\mathbf{s}$. Furthermore, we assume that there is time synchronization in the system for coherent uplink transmissions.

A. Channel model

We consider a uniform linear array (ULA) whose response vector can be expressed as

$$\mathbf{a}(\omega) = [1 \quad e^{-j\omega} \quad \dots \quad e^{-j(M-1)\omega}]^T \tag{5}$$

where $\omega = \frac{2\pi d \sin\theta}{\lambda}$, d is the antenna spacing at the base station, λ is the signal wavelength and θ is angle of arrival of single path. The received signal at the base station can be expressed as a multipath model utilizing the response array vector as

$$\mathbf{h}_{lc} = \sum_{i=1}^{Q} \gamma_i \mathbf{a}(\omega_i) \tag{6}$$

where γ_i is complex random gain factor, θ_i is the angle of the arrival of the ith path, Q is the number of paths. We adopt a generic Toeplitz correlation model as it is the suitable model for implementing the correlation from theoretical [25] and practical perspectives [29]. The two generic correlation types have a generic Toeplitz structure.

In order to analytically study the performance of the proposed technique, we deal with a simplified exponential correlation model due to its mathematical tractability [25]. The correlation structure of $\mathbf{R}_{lc}$ can be formulated as following

1) Exponential Correlation [25]: It is known that exponential correlation matrix is a special case of Toeplitz and it is often used for ULA system, and it has the following formulation

$$[\mathbf{R}(i,j)] = \begin{cases} \rho^{|i-j|}, i > j \\ \rho^{|i-j|*}, i < j \end{cases} \tag{7}$$

where $\rho \in \mathbb{C}, |\rho| \leq 1$. This kind of correlation is suitable for theoretical analysis, and it will be used in the next sections to study the benefits of the proposed framework.

2) Practical Correlation [29]: In order to approximate the practical correlation, the received signal at the base station can be implemented as limited memoryless multipath model with single tap utilizing the response array vector as (6). For a multipath scattering confined to a relatively small angular spread seen from the base station. A general correlation structure can be well approximated by

$$\mathbf{R} \approx \mathbf{D}_a \mathbf{B} \mathbf{D}_a^H$$

where $\sigma_\omega = 2\pi \frac{d}{\lambda}\sigma_\theta \cos\theta$, $\mathbf{D}_a = diag[\mathbf{a}(\omega)]$. σ_θ is the standard deviation of the angular spread.

$\mathbf{B}$ depends on the angular spread of the multipath components. The angular distribution is Gaussian $\tilde{\omega} \in \mathcal{N}(0, \sigma_\omega)$, it can be written as

$$[\mathbf{B}(m,n)] \simeq \exp(\frac{((m-n)\sqrt{3}\delta_\omega)^2}{2}). \tag{8}$$

For $\tilde{\omega}$ uniformly distributed over $[-\delta_\omega, \delta_\omega]$, it has the following structure

$$[\mathbf{B}(m,n)] \simeq \frac{\sin((m-n)\delta_\omega)}{(m-n)\delta_\omega}. \tag{9}$$

and $\sigma_\omega = \sqrt{3}\delta_\omega$. We adopt the both correlation models: exponential and practical to test the efficiency of the proposed algorithms. From the previous correlation expressions, it can be argued that each user in the cell has a different covariance

matrix due to its position as the covariance matrix is a function of the angular spread and its corresponding distribution.

The covariance information of the target users and interfering users can be acquired exploiting resource blocks where the desired user and interference users are known to be assigned training sequences at different times. Alternatively, this information can be obtained using the knowledge of the approximate users' positions and the type of the angular spread at BS side exploiting the correlation equations (8)-(9).

III. DCT FOR SPATIAL COMPRESSION

The optimal transform that decorrelates the signal is Karhunen-Loéve Transform (KLT) which requires significant computational resources [27]. Therefore, fixed transforms are preferred in many applications. In contrast to data-dependent transform KLT, DCT is a fixed transform that does not depend on the data structure and has excellent energy compaction properties that perform very close to KLT. Although there are several fixed transform techniques (i.e. FFT, DFT), DCT is more useful in the context of this paper since it has higher compression efficiency in comparison with the other techniques. DCT is a technique for converting a signal into elementary frequency components. It transforms the signal from time domain to frequency domain. Most of the signal information tends to be concentrated in a few components of the DCT if the information is correlated [26]. Therefore, the signal can be compressed by keeping the important frequencies and truncating the least influential ones without impacting the quality of estimate. DCT's qualities motivate its utilization in time-domain estimation in OFDM-MIMO to reduce the border effect owing to its capacity to reduce the high frequency components in the transform domain [28]. In this work, we do not tackle the multicarrier OFDM, we utilize the DCT to handle the interference at the estimation process in multiple antenna multicell systems.

The compression capability of DCT[1] is depicted in Fig. 1 for a covariance matrix as in (7) with $\rho = 0.9$. The signal energy is condensed in the first few spatial frequencies, and it can be noticed that 93% of signal is compressed in the first four coefficients. The DCT capability compression helps in categorizing the important and the unimportant spatial frequencies in order to concentrate the contamination in fewer dimensions. In order to evaluate the effectiveness of DCT compaction property, we study the correlation models mentioned in the system model section.

In order to study numerically the compression capability of DCT, we note that $\mathbf{R}$ spans a set of eigenspaces which determines the directions of the transmissions. Each eigenspace acquires certain power, which is known as the eigenvalue of this space. In this direction, we use the metric of the standard condition number (SCN), which is defined as the ratio of the maximum eigenvalue to the minimum eigenvalue and denoted by $\chi(\mathbf{R})$, before and after DCT has been applied. The intuition behind using the SCN is to measure the dynamic

[1]See eq. (12)- (14) for more details about DCT.

range of eigenvalues which is representative of the compression level. This can be demonstrated the reduction of SCN, since preserving a small number of strong eigenvalues means that the signal space can be effectively represented by fewer eigenvectors. Thus, the same energy amount is contained in lower dimensional subspace which indicates a compression. The maximum and the minimum eigenvalues for exponential correlation matrix $\mathbf{R}$ as in eq. (7) are $\frac{1}{(1-\rho)^2}$ and $\frac{1}{(1+\rho)^2}$ respectively [34].

Lemma 1. *[34]: The SCN of any exponential-form correlation matrix $\mathbf{R}$ is thus given by*

$$\lim_{M \to \infty} \chi(\mathbf{R}) = \left(\frac{1+\rho}{1-\rho}\right)^2 \tag{10}$$

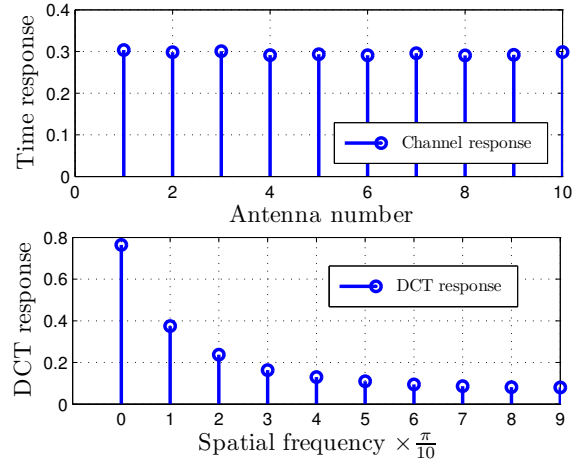*The SCN can be very large for highly correlated data ($\rho \to 1$).*

Fig. 1. *Comparison between the DCT spatial frequency response and spatial time response at the scenario of $\rho = 0.9$ in (7).*

The DCT basis comprises of the eigenvectors of the following symmetric tri-diagonal matrix [33]

$$\mathbf{Q}_c = \begin{bmatrix} 1-\zeta & -\zeta & 0 & \dots & 0 \\ -\zeta & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & -\zeta & 1 & -\zeta \\ 0 & \dots & 0 & -\zeta & 1-\zeta \end{bmatrix} \tag{11}$$

where $\zeta = \frac{\rho}{1+\rho^2}$. In matrix form, the eigenvectors are given by

$$[\mathbf{U}_D(k,n)] = c[k] \cos\left(\frac{(2n+1)k\pi}{2M}\right). \tag{12}$$

As a result, the definition of DCT has the following interpretation

$$\mathbf{m}_d[k] = c[k] \sum_{n=0}^{M-1} \mathbf{m}[n] \cos\left(\frac{\pi(2n+1)k}{2M}\right), \tag{13}$$

where $\mathbf{m}$ is the time domain vector and $\mathbf{m}_d$ is the DCT representation of $\mathbf{m}$ and can be written as:

$$\mathbf{m}[n] = \sum_{k=0}^{M-1} c[k]\mathbf{m}_d[k] \cos\left(\frac{\pi(2n+1)k}{2M}\right), \qquad (14)$$

where k represents the spatial frequency coefficient, n denotes the antenna number, and finally $c[k]$ is defined as below:

$$c[k] = \begin{cases} \sqrt{\frac{1}{M}} & , k = 0, \\ \sqrt{\frac{2}{M}} & , \text{otherwise}. \end{cases}$$

A. DCT for Uniform Linear Array

In order to interpret the DCT channel response, we need to find the DCT for ULA response, which can be expressed as (15). From (15), the DCT for channel vector following the expression in (6) can be implemented as (16). To verify the compression efficiency of employing DCT, we study its impact on the change of eigenvalues of the exponential correlation matrix.

Lemma 2. *[34] The SCN of the exponential correlation matrix eq.(7) after the DCT is such that*

$$\lim_{M \to \infty} \chi(\mathbf{U}_D \mathbf{R} \mathbf{U}_D{}^H) = 1 + \rho. \qquad (17)$$

The change of SCN from $\frac{1+\rho}{1-\rho}$ to $1+\rho$ before and after DCT transformation explains the compressive nature of DCT.

The correlation matrix can be re-implemented as $\mathbf{T}_{lc} = \mathbf{E}[\mathbf{m}_{lc}^d \mathbf{m}_{lc}^d{}^H] = \mathbf{A}_D \mathbf{R}_{lc} \mathbf{A}_D^H$, which is a 2-dimensional DCT of the $\mathbf{R}_{lc}$. Decontaminating the training sequences from the interference requires taking into the account to the following important points

- As most information is condensed in the certain frequencies, these frequencies should face the lowest possible interference in order to obtain an accurate estimate. The spatial frequencies are functions of angles of arrival of pilots at the BS, thus condensing these spatial frequencies in certain bands makes the separation process easier. This kind of separation can be enhanced by prearranging the spatial frequencies through performing training sequence allocation which is discussed later in section VII.

- The users who have the least common spatial characteristics should be assigned the same sequence; as they have the minimum overlap in the spatial frequency domain.

The compression pattern of DCT is defined as a function of the power distribution among the spatial frequencies. Such distributions depend on the nature of the angular spread at the BS. To enable the spatial separation among the different estimate, the pattern of compression should be highlighted. For this purpose, we adopt the correlation models of [8] to formulate generic rules about DCT compression for ULA systems.

Lemma 3. *The spatial DCT frequencies are concentrated at low frequencies if angle of arrivals are close to zero. If the direction of the arrivals are close to $\frac{\pi}{2}$, the DCT components are condensed at high frequencies.*

Proof. See Appendix. $\qquad \square$

The previous Lemma 3 confirms the intuition that the contamination is minimum when the important frequencies of the users who allocate the same training sequence are concentrated at different spatial band. The compression efficiency enhances with increasing the number of antennas at the BS, which makes the separation much easier.

Lemma 4. *As $M \to \infty$, the DCT response is condensed in a single spatial frequency component.*

Proof. See Appendix. $\qquad \square$

In the next section, we utilize the DCT compression to deal with the problem of contaminated estimation in correlated multiuser environment.

IV. CHANNEL ESTIMATION WITH $\mathcal{K}$- TRAINING SEQUENCE REUSE

Exploiting the ULA structure, we develop a new estimator with the aim of decontaminating the reused training sequences over the network. Our estimators utilize the embedded information in the second order statistics of the channel vectors. The covariance matrices capture the embedded information related to the distribution (mainly mean and spread) of the multi-path angles of arrival at the base station [31]. In [10]-[11], the authors focus on optimal training sequence designs and they exploit the covariance matrices of the desired channels and interference. The optimal training sequences are developed with adaptation to the statistics of the disturbance [10]. An extension of [10] is proposed in [30], which proposes a more general framework for the purpose of training sequence design in MIMO systems, which handles not only minimization of channel estimators MSE as an optimization metric, but also the optimization of a final performance metric of interest related to the use of the channel estimate in the communication system. However, the design of the training sequence does not have an impact on interference mitigation, as long as, we utilize fully aligned training sequences. Here, we concentrate on designing an estimation technique that can achieve accurate results, by exploiting the spatial frequency to mitigate the interference in CSI estimation process.

A. Bayesian Estimation

The Bayesian Estimator (BE) is widely discussed in the literature and is utilized in many applications [32]. The BE coincides with the minimum mean square estimator in case $\mathbf{h}_{lc}$ and $\mathbf{y}_c$ are jointly Gaussian distributed random variables. The BE estimator can be formulated as [32]

$$\mathbf{F}_{lc} = \mathbf{G}_{lc}\mathbf{S}^H = \mathbf{R}_{lc}\left(\left(\sum_{l=1}^{C} \mathbf{R}_{lc}\right) + \frac{\sigma_n{}^2}{\tau}\mathbf{I}_M\right)^{-1}\mathbf{S}^H. \qquad (18)$$

The previous formulation provides insights about the nature of the training sequence contamination; the estimator is a function of all the correlation matrices related to all users who utilize the same sequence. Therefore, the spatial characteristics of all users, who have the same sequence, influence on the accuracy of the estimations.

$$
\mathbf{a}^{\text{DCT}}[k] = \begin{cases} \sqrt{\frac{1}{M}} e^{-j\omega_i \frac{M-1}{2}} \frac{\sin\frac{M}{2}\omega_i}{\sin\frac{\omega_i}{2}} & , k = 0, \\ \sqrt{\frac{2}{M}} \frac{\sin((\omega_i+\frac{\pi k}{M})\frac{M}{2})}{\sin((\omega_i+\frac{\pi k}{M})\frac{1}{2})} e^{j\frac{(\omega_i+\frac{\pi k}{M})(M-1)}{2}+\frac{\pi k}{M}} + \sqrt{\frac{2}{M}} \frac{\sin((\omega_i-\frac{\pi k}{M})\frac{M}{2})}{\sin((\omega_i-\frac{\pi k}{M})\frac{1}{2})} e^{j\frac{(\omega_i-\frac{\pi k}{M})(M-1)}{2}-\frac{\pi k}{M}} & , 0 \le k \le M-1. \end{cases} \tag{15}
$$

$$
\mathbf{h}_{lc}^{\text{DCT}}[k] = \begin{cases} \sum_{i=1}^{P} \frac{\gamma_i \sqrt{\frac{1}{M}}}{\sqrt{P}} e^{-j\omega_i \frac{M-1}{2}} \frac{\sin\frac{M}{2}\omega_i}{\sin\frac{\omega_i}{2}} & , k = 0, \\ \sum_{i=1}^{P} \frac{\gamma_i \sqrt{\frac{2}{M}}}{\sqrt{P}} \left(\frac{\sin((\omega_i+\frac{\pi k}{M})\frac{M}{2})}{\sin((\omega_i+\frac{\pi k}{M})\frac{1}{2})} e^{j\frac{(\omega_i+\frac{\pi k}{M})(M-1)}{2}+\frac{\pi k}{M}} + \frac{\sin((\omega_i-\frac{\pi k}{M})\frac{M}{2})}{\sin((\omega_i-\frac{\pi k}{M})\frac{1}{2})} e^{j\frac{(\omega_i-\frac{\pi k}{M})(M-1)}{2}-\frac{\pi k}{M}} \right) & , 0 \le k \le M-1. \end{cases} \tag{16}
$$

1) Minimum Mean Square Error Performance : The considered performance is the mean square error (MSE) of the proposed estimator, and can be expressed as [32]

$$
\mathcal{E}_{lc} = \mathbb{E}_{\mathbf{h}_{lc}} \left\{ \|\hat{\mathbf{h}}_{lc} - \mathbf{h}_{lc}\|^2 |\hat{\mathbf{h}}_{lc} \right\}. \tag{19}
$$

The final formulation for MSE can be expressed as

$$
\mathcal{E}_{lc} = tr\left(\mathbf{R}_{lc} - \mathbf{R}_{lc}^2 \Big(\sum_{m=1}^{C} \mathbf{R}_{mc} + \frac{\sigma_n^2}{\tau} \mathbf{I}_M \Big)^{-1} \right). \tag{20}
$$

The upper bound of the MSE of the BE can be written as

$$
\mathcal{E}_{lc}^{MI} = tr\left(\mathbf{W}_{lc} \Big(C\mathbf{\Delta}_{lc} + \frac{\sigma_n^2}{\tau} \mathbf{I}_M \Big)^{-1} \mathbf{W}_{lc}^H \right) \tag{21}
$$

where $\mathbf{R}_{lc}$ is decomposed as $\mathbf{W}_{lc} \mathbf{\Delta}_{lc} \mathbf{W}_{lc}^H$, and the subscript MI refers to the "maximum interference scenario". This scenario occurs when the set of users, who utilizes the same training sequence, has identical spatial second order statistic and attenuation towards certain BS.

The lower bound of the MSE of the BE is as:

$$
\mathcal{E}_{lc}^{NI} = tr\left(\mathbf{R}_{lc} \Big(\mathbf{R}_{lc} + \frac{\sigma_n^2}{\tau} \mathbf{I}_M \Big)^{-1} \right) \tag{22}
$$

where superscript NI refers to the "no interference scenario". This lower bound can be achieved when the users span distinct subspaces and this condition should be satisfied $\{\mathbf{W}_{lc}\mathbf{W}_{lj} = \mathbf{0}, \forall j \neq c\}$.

Therefore, the overlap in these subspaces will degrade the estimate. A new look to the problem will be handled through the DCT framework in the next sections. As the work in this paper aims at minimizing the estimation errors to reduce the contamination, we do not consider beamforming which is handled in [14]. Taking this into the account, we only consider conventional beamforming techniques:

- Coordinated Beamforming (CB) requires the estimation of the intefering channels. The contamination occurs when the same BS assigns the same sequence to estimate the served user channel as well as interfering channels and/or other BSs use the same training to estimate their users' channel or the corresponding interfering channels. In this scenario, the contamination can be utilized to estimate the interfering channel for the user that utilizes the same training sequence.
- Maximum ratio transmission (MRT), which requires the estimation of the desired user channel. The contamination

happens when the base stations utilize the same training sequence for their users.

B. Least Square Estimation

Least square estimation can be utilized in the scenarios when the information about the second order statistics is not available. Hence, if the received signal is modeled as (2), a least square (LS) estimator for the desired channel can be formulated as

$$
\hat{\mathbf{h}}_{lc}^{\text{ls}} = \mathbf{S}^H \mathbf{y}_c. \tag{23}
$$

The conventional estimator suffers from a lack of orthogonality between the desired and interfering training sequences, an effect known as training sequence contamination [14], [16]-[17]. In particular, when the same training sequence is reused in all cells, the estimated channel can be expressed as

$$
\hat{\mathbf{h}}_{lc}^{\text{ls}} = \mathbf{h}_{lc} + \sum_{m \neq c}^{C} \mathbf{h}_{mc} + \frac{\mathbf{S}^H \mathbf{N}}{\tau}. \tag{24}
$$

As it appears in (24), the interfering channels have strong impact and leak directly into the desired channel estimate. The estimation performance is then limited by the signal to interfering ratio at the base station, which consequently limits the ability to design effective beamforming solutions.

1) Mean Square Error Performance: The MSE for LS can derived as (19), and the closed form expression can be formulated as

$$
\mathcal{E}_{lc}^{LS} = tr\left(\sum_{m=1, m \neq l}^{C} \mathbf{R}_{mc} \right). \tag{25}
$$

From (25), it can be noted that MSE is a linear function of the involved correlation matrices traces, and it grows linearly with the number of users who utilizes the same training sequence.

V. ESTIMATION TECHNIQUES USING DCT COMPRESSION

A. Important Frequencies Determination

In order to decontaminate the channel estimate, we need to determine the important frequencies that should be extracted from the estimated channel $\mathbf{S}^H \mathbf{y}_c$. To determine these important frequencies, we need to solve the following quadratic form

$$
\omega_i^* = \arg\max_{\omega_i^*} \quad \mathbf{a}_{\text{D}}(\omega_i) \mathbf{R}_{lc} \mathbf{a}_{\text{D}}^H(\omega_i). \tag{26}
$$

The solution lies in finding the eigenvectors related to the maximum eigenvalues. Therefore, the concentration of the

important spatial frequencies depend on the angular spread at the BS. In order to separate the contamination from the required estimate, we need to determine the number of the important frequencies. It should be noted that the number of considered spatial frequency depends on the tradeoff between the loss resulted from contamination and DCT lossy compression. We define the vector $\mathbf{q}_{lc} \in \{0,1\}^{M \times 1}$, which identifies the unimportant and important frequencies of $\mathbf{h}_{lc}$ as 0 and 1 respectively. A definition of the compression ratio of can be stated as

$$\eta = \frac{\|\mathbf{q}_{lc}\|_0}{M}. \tag{27}$$

In the next section, integrated DCT-LS and DCT-BE frameworks are proposed to deal with the problem of contaminated estimation in correlated multiuser environment.

B. DCT Based Bayesian Estimation (DBE)

As the received signal is a combination of all users' channel who utilize the training sequence, the DCT of the received signal contains the spatial frequencies related to these channels compressed in certain components. Therefore, a splitting technique in DCT domain is applicable if we know the most important frequencies related to the required estimate. To reduce the impact of contamination, we can extract these frequencies and replace the least important frequencies by their average values, which are close to zero. This can be explained by Fig.(2)-a (the upper figure), it depicts the splitting capability of DCT by distinguishing the important information of the involved estimate, which is clearly condensed at high and low frequencies. For the first estimate (the black one), the information is compressed in the low spatial frequency components while the high spatial frequencies do not hold much information. It can be noticed that these spatial frequencies are contaminated by other channels (the blue and the pink ones). This contamination can be tackled by removing these spatial frequencies and equating them to zero.
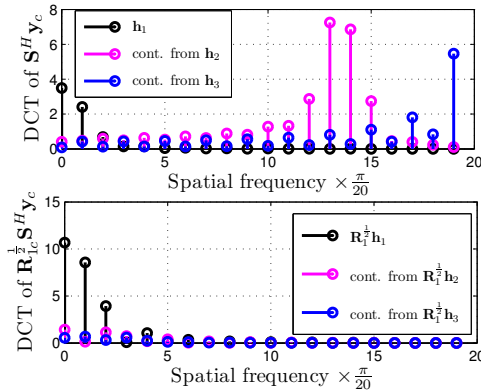

Fig. 2. *A comparison of DCT response employing before and after multiplication by $\mathbf{R}_{lc}^{\frac{1}{2}}$ 'cont' denotes contamination.*

The new estimation technique should take into consideration the covariance information and DCT compression capability to boost the performance of BE. The new estimation procedure can be described as following

---

A1. DCT based Bayesian Estimation (DBE)

- At user's side: Send the training sequence $\mathbf{s}$.
- At BS side:
  1) Modify the covariance matrices in $\mathbf{G}_{lc}$ using $\hat{\mathbf{R}}_{lc}$, which can be formulated as follows:

$$\begin{aligned} \hat{\mathbf{U}}_D &= \mathbf{U}_D \odot \left(\mathbf{q}_{lc} \otimes \mathbf{1}_{1 \times M}\right) \\ \mathbf{U}_D \hat{\mathbf{R}}_{lc} \mathbf{U}_D^H &= \hat{\mathbf{U}}_D \mathbf{R}_{lc} \hat{\mathbf{U}}_D^H \end{aligned}$$

  2) Find $\hat{\mathbf{y}}_c = \mathbf{S}^H \mathbf{y}_c$, extract the important information related to the estimate $\tilde{\mathbf{y}}_c = (\mathbf{U}_D \hat{\mathbf{y}}_c) \odot \mathbf{q}_{lc}$.
  3) Take the inverse DCT of $\tilde{\mathbf{y}}_c$ as $\mathbf{U}_D^H \tilde{\mathbf{y}}_c = \mathbf{y}_c^f$.
  4) Find $\hat{\mathbf{h}}_{lc} = \mathbf{G}_{lc} \mathbf{y}_c^f$.

C. DCT Based Least Squares Estimation (DLS)

In comparison with typical LS, in which the estimation results in direct summation of all channel (24), the DCT Based LS (DLS) requires information about the important frequency set for each estimated channel. The estimation can be summarized as follows

---

A2. DCT based Least Square Estimation (DLS)

- At the user's terminals: Send the training sequence $\mathbf{s}$
- At the BS's side:
  1) Employ typical LS on the received signal, find $\hat{\mathbf{h}}_{lc}^{ls}$.
  2) Employ DCT on $\hat{\mathbf{h}}_{lc}^{ls}$, extract the important DCT frequencies related to estimate by $\tilde{\mathbf{h}}_{lc}^{ls} = (\mathbf{U}_D \hat{\mathbf{h}}_{lc}^{ls}) \odot \mathbf{q}_{lc}$.
  3) Take the inverse DCT of the extracted version $\mathbf{U}_D^H \tilde{\mathbf{h}}_{lc}^{ls}$.

This estimation technique can be combined with a training sequence allocation algorithm to make the most important spatial frequencies distinct which simplifies the separation of these frequencies. The DLS can be used to estimate the channel for MRT beamforming since it only requires the statistical information of the target channel.

The major difference between DBE and DLS is that DBE decontaminates the training sequences using two steps: zeroing the unimportant frequencies, and the BE step to remove the contamination from the important frequencies. BE and DBE require the covariance acquisition of the direct and all interfering links. Another complexity source is the matrix inversion. Thereby, their usage in large scale MIMO systems is limited and constrained to the limited size problems. On the other hand, DLS only requires the covariance knowledge of the target channel. It has less complexity due to the absence of a matrix inversion step and the acquisition of the interfering links' covariance matrices.

VI. MODIFIED SPATIAL ESTIMATION

To enhance the performance of BE, DBE and DLS, the important frequencies should be more distinguishable. This can be obtained by compressing them into smallest possible number of components to separate them more effectively. Therefore, a modification can be proposed to handle this issue using the available correlation information and DCT.

The concept of increasing the correlation of the channel of interest is proposed in this section. This makes the target channel more ill-conditioned and thereby the contamination

more separable. Therefore, to estimate $\mathbf{h}_{lc}$ we can increase its correlation by multiplying the received signal $\mathbf{y}_c$ in (4) by $\mathbf{R}_{lc}^{\frac{i}{2}}\mathbf{S}^H$, which makes the correlation of the target estimate $\mathbf{R}_{lc}^{i+1}$ and the correlation of contamination $\mathbf{R}_{lc}^{\frac{i}{2}}\mathbf{R}_{mc}\mathbf{R}_{lc}^{\frac{i}{2}}$.

The effect of implementing such step for $i = 1$ is depicted in Fig.(2)-b. In the considered scenario, $\theta_s = \{0°, 15°, 35°\}$ are the lower limit of the angular spread for each channel respectively, $\Delta\theta = 20°$ is the angular spread. It can be noted that after employing this step, the impact of contamination in the DCT domain is limited and does not have any influential contribution to the estimate and can be removed easily using DCT. Taking into the account this property, we propose two estimation techniques as follows
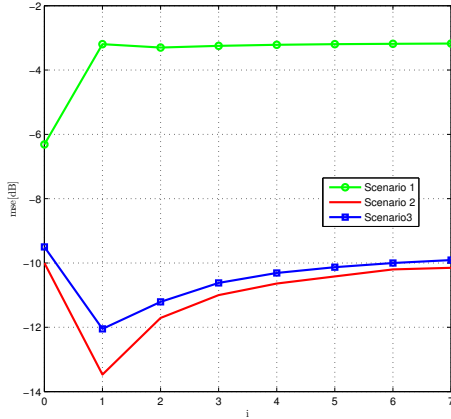

Fig. 3. *The normalized MSE vs. $\mathbf{R}_{lc}^{\frac{i}{2}}$ at different correlation scenarios, $K = 3$, $\Delta\theta = 20°$. Scenario 1:$\theta_s = [0°, 20°, 40°]$, Scenario 2:$\theta_s = [0°, 15°, 30°]$, Scenario 3:$\theta = [0°, 10°, 20°^{3irc}]$*

A. Modified BE Based Estimation (MBE)

The multiplication by $\mathbf{R}_{lc}^{\frac{i}{2}}$ increases the correlation of the target estimate, the modified correlation can be decomposed as $\mathbf{W}_{lc}\mathbf{\Delta}_{lc}^{i+1}\mathbf{W}_{lc}^H$. This makes eigenvalues of the target estimate more distinct since the stronger eigenvalue becomes more effective and the opposite holds for the weaker ones. Therefore, the channel becomes more ill-conditioned and more separable. However, the impact of this multiplication is unknown with respect to the interfering subspaces and heuristic approach is proposed to gain the benefit of it when it is applicable. After the multiplication, the received signal can be written as

$$\mathbf{y}_c^{'} = \gamma_{lc}\mathbf{R}_{lc}^{\frac{i}{2}}\mathbf{S}^H\mathbf{y}_c = \gamma_{lc}\mathbf{R}_{lc}^{\frac{i}{2}}\mathbf{S}^H\mathbf{S}\sum_{m=1}^{C}\mathbf{h}_{lc} + \mathbf{R}_{lc}^{\frac{i}{2}}\mathbf{S}^H\mathbf{n}_c \quad (28)$$

where γ_{lc} is designed to keep the signal power fixed $\|\mathbf{y}_c^{'}\|^2 = \|\mathbf{y}_c\|^2$. The modified BE can be derived in a similar fashion to traditional BE (19)-(23) in the original manuscript but with assuming that correlation matrices $\mathbf{Q}_{mc} = \mathbf{R}_{lc}^{\frac{i}{2}}\mathbf{R}_{mc}\mathbf{R}_{lc}^{\frac{i}{2}}$. To utilize the enhanced compression capability, multiply the received signal by $\mathbf{R}_{lc}^{\frac{i}{2}}\mathbf{S}^H$ and then employ the following filter

to estimate the time domain $\mathbf{h}_{lc}$[2]. This makes the formulation of the final estimation at the BS as

$$\mathbf{P}_{lc}^i = \gamma_{lc}\mathbf{R}_{lc}^{1+\frac{i}{2}}\left(\mathbf{R}_{lc}^{\frac{i}{2}}\left(\sum_{m=1}^{C}\mathbf{R}_{mc}\right)\mathbf{R}_{lc}^{\frac{i}{2}} + \frac{\sigma_n^2}{\tau}\mathbf{R}_{lc}^i\right)^{-1}\mathbf{R}_{lc}^{\frac{i}{2}}\mathbf{S}^H \quad (29)$$

The total MSE of estimating $\mathbf{h}_{lc}$ can be evaluated as

$$\epsilon_{lc}^i = \mathbb{E}_{\mathbf{h}_{lc}}\left\{tr\left((\mathbf{h}_{lc} - \mathbf{P}_{lc}^i\mathbf{y}_c)(\mathbf{h}_{lc} - \mathbf{P}_{lc}^i\mathbf{y}_c)^H\right)\right\} \quad (30)$$

and finally the MSE can have the following closed form expression

$$\epsilon_{lc}^i = tr\left(\mathbf{R}_{lc} - \gamma_{lc}^2\mathbf{R}_{lc}^{i+2}\left(\mathbf{R}_{lc}^{\frac{i}{2}}(\sum_{m=1}^{C}\mathbf{R}_{mc})\mathbf{R}_{lc}^{\frac{i}{2}} + \frac{\sigma_n^2}{\tau}\mathbf{R}_{lc}^i\right)^{-1}\right). \quad (31)$$

To study the impact of this multiplication on the MSE of different MBE is studied in Fig.(3), in which the MSE is depicted with respect to i, where $i = 0$ refers to typical BE. It can be concluded at scenario 2 and 3 that this multiplication reduces the MSE, which motivates the utilization of this step at the estimation process. On contrary to the two scenarios, the first one shows a contradicting performance. This can be explained that without the multiplication, it is impossible to separate the channel estimations due to the complete overlap among the users' subspaces. This factor passively enhances in case of this multiplication due to the noise amplification factors. Therefore, an adaption between the typical BE and modified BE is required to get the joint benefits of the both techniques. Moreover, it can be noted that the best variant (the ith power) of $\mathbf{R}_{lc}$ from the MSE perspective is the first one and it is sufficient to be used in the estimation.

In order to optimize the system performance, we can adapt the estimation strategy based on the involved users who utilize the same training sequence. The adaptive strategy $\mathcal{J}_l$ can be expressed as

$$\mathcal{J}_l = \begin{cases} BE & , \mathcal{E}_{lc}^{BE} \leq \mathcal{E}_{lc}^{MBE} \\ MBE & , \mathcal{E}_{lc}^{BE} \geq \mathcal{E}_{lc}^{MBE}. \end{cases} \quad (32)$$

The selection procedure is implemented at the system installation level and is performed once for each channel. Therefore, the users who utilize the same training sequence can have different estimation techniques. It can be noted that the important frequencies of the target estimate do not change with the multiplication of $\mathbf{R}_{lc}^{\frac{1}{2}}$, since the eigenvectors of the estimate are not change by this multiplication. However, the subspaces spanned by interfering signals change due to the modified covariance matrices $\hat{\mathbf{R}}_{lm} = \mathbf{R}_{lc}^{\frac{1}{2}}\mathbf{R}_{lm}\mathbf{R}_{lc}^{\frac{1}{2}}$. However, there is no mathematical explanation clarifies the relation between the eigenvectors of $\mathbf{R}_{lm}$ and $\hat{\mathbf{R}}_{lm}$. This makes the selection of the most appropriate estimation technique hard, thus the adaptation based on MSE is the solution for this issue. The adaptation depends on the subspaces spanned by the target and interfering signals related to traditional and modified techniques. Moreover, the multiplication by other variants can lead to excessive contamination correlation with

[2]It is derived in a similar fashion to Bayesian estimation technique using different correlation set $\mathbf{R}_{lc}^m = \mathbf{R}_{lc}^2$, the estimated channel for the cth user and $\mathbf{R}_{lm}^m = \mathbf{R}_{lc}^{\frac{1}{2}}\mathbf{R}_{lm}\mathbf{R}_{lc}^{\frac{1}{2}}$.

target subspaces, this makes the contamination subspaces span the same subspaces as the target estimate which actually makes the contamination mitigation much harder.

B. Modified DCT Bayesian Estimation (MDBE)

The concept of integrating the DCT with modified BE can be adopted to enhance the estimation quality, the modified DCT BE can be described as following

---
A3. Modified DCT Bayesian estimation (MDBE)

- Multiply $\mathbf{R}_{lc}^{\frac{1}{2}}\mathbf{S}^H$ by the received signal $\mathbf{y}_c$ and get $\mathbf{g}_{lc}$.
- Employ the same steps (1)-(4) as DBE at $\mathbf{g}_{lc}$ to $\hat{\mathbf{g}}_{lc}$. It should be noted that $\mathbf{G}_{lc}$ should be replaced $\mathbf{P}_{lc}$.
- To obtain the estimate $\hat{\mathbf{h}}_{lc} = \hat{\mathbf{g}}_{lc}$.

The adaptation idea can be exploited as in (32) to enhance the performance of DCT based estimation to optimally select the estimation technique from DBE or MDBE.

C. Modified DCT LS Estimation (MDLS)

The joint utilization of the DCT and LS within the modified framework is proposed to boost the quality of LS estimate, and can be summarized as follows

---
A4. Modified DCT LS estimation (MDLS)

- Multiply $\mathbf{R}_{lc}^{\frac{1}{2}}\mathbf{S}^H$ by the received signal $\mathbf{y}_c$ and get $\mathbf{g}_{lc}$.
- Employ the same steps (1)-(4) as DLS at $\mathbf{g}_{lc}$ to $\hat{\mathbf{g}}_{lc}$.
- To obtain the estimate $\hat{\mathbf{h}}_{lc} = \mathbf{R}_{lc}^{-\frac{1}{2}}\hat{\mathbf{g}}_{lc}$.

---

rrelation matrix structure $\mathbf{R}_{lc} = \mathbf{W}_{lc}\boldsymbol{\Delta}_{lc}\mathbf{W}_{lc}^H = \hat{\mathbf{W}}_{lc}\hat{\boldsymbol{\Delta}}_{lc}\hat{\mathbf{W}}_{lc}$, where $\hat{\boldsymbol{\Delta}}_{lc} \in \mathbb{R}_+^{r \times r}, \hat{\mathbf{W}}_{lc} \in \mathbb{C}^{M \times r}$ are the matrices that contain the positive eigenvalues and the eigenvectors that are associated with the positive eigenvalues. This enables the implementation of the proposed algorithms in the case of ill-conditioned channels. The usage of DLS and MDLS can be adapted like (32) to enhance the quality of estimation.

VII. JOINT SPATIAL ESTIMATION AND TRAINING SEQUENCES ALLOCATION

Several training sequence assignment techniques are investigated in the recent literature [21]- [22], in order to select the optimal set of users that can utilize the same training sequence simultaneously. These allocation techniques mimic the previously proposed scheduling algorithms in multiuser MIMO scenarios [5] as they try to assign the same training sequence to the users who have distinct subspaces. The training sequence assignment strategies inherit the concept of multiuser diversity, which depends on the pool of the involved users, as a consequence, the system performance is degraded for low number of users. In this section, an heuristic approach assigns the available training sequences to required users' channel estimates is proposed as

A. Training Sequence Allocation for BE related techniques

Define the set of users who utilizing the training sequence $\mathbf{s}_l$ by $\mathcal{U}(\mathbf{s}_l)$, and the set of user who adopt the modified BE estimation as $\mathcal{U}_M(\mathbf{s}_l) \subset \mathcal{U}(\mathbf{s}_l)$ and the set of users whose channel are estimated by typical BE $\mathcal{U}_B(\mathbf{s}_l) \subset \mathcal{U}(\mathbf{s}_l)$ such that $\mathcal{U}_M(\mathbf{s}_l) \cup \mathcal{U}_B(\mathbf{s}_l) = \mathcal{U}(\mathbf{s}_l)$ and the set of all users $\mathcal{U}$. The set of all users who have not assigned a training sequence yet is defined as $\mathcal{M}$. We define a measuring function as (33), the allocation algorithm is described as

---
A5. Greedy Training Sequence Allocation Algorithm for (BE/MBE) estimation

---

- Start from any training sequence, without any loss of generality we start with $\mathbf{s}_1$.
- Determine the training sequence reuse factor $\mathcal{K}$.
- Define the set of users who should allocate the same training sequence as $\mathcal{U}(\mathbf{s}_l)$.
- Pick random user $u \in \mathcal{M} = \mathcal{U} - \cup_{n=1}^{l-1}\mathcal{U}(\mathbf{s}_n)$
- Step 1: set $\mathcal{U}(\mathbf{s}_l) = u$
- Step 2: if $|\mathcal{U}(\mathbf{s}_l)| \leq \mathcal{K}$
  1) for users= $\mathcal{M}(1)$ to $\mathcal{M}(|\mathcal{M}|)$

$$u^* = \min_u \Big(\min_{u_M \in \mathcal{M}} \mathcal{E}(\mathcal{U}_B(\mathbf{s}_l), \mathcal{U}_M(\mathbf{s}_l) \cup \{u_M\})$$
$$, \min_{u_B \in \mathcal{M}} \mathcal{E}(\mathcal{U}_B(\mathbf{s}_l) \cup \{u_B\}, \mathcal{U}_M(\mathbf{s}_l)) \Big)$$

  2) $\mathcal{M} = \mathcal{M}/\{u^*\}$
  3) Estimation technique selection for u^* is done according to

$$\begin{cases} \text{MDBE}, \mathcal{E}(\mathcal{U}_B(\mathbf{s}_l), \mathcal{U}_M(\mathbf{s}_l) \cup \{u^*\}) \leq \mathcal{E}(\mathcal{U}_B(\mathbf{s}_l) \cup \{u^*\}, \mathcal{U}_M(\mathbf{s}_l)) \\ \text{DBE}, \mathcal{E}(\mathcal{U}_B(\mathbf{s}_l), \mathcal{U}_M(\mathbf{s}_l) \cup \{u^*\}) \leq \mathcal{E}(\mathcal{U}_B(\mathbf{s}_l) \cup \{u^*\}, \mathcal{U}_M(\mathbf{s}_l)) \end{cases}$$

  4) $\mathcal{U}(\mathbf{s}_l) = \mathcal{U}(\mathbf{s}_l) \cup \{u^*\}$
  5) go to step 2.

This algorithm aims at minimizing the estimation error in two steps: the first step is summarized by allocating training sequences to the set of users whose spatial signatures have the most distinctive characteristics. The performance improves with the number of user as it becomes more likely to find users with distinct second order statistics to be assigned the same training sequence. The second step is the selection of the optimal estimation technique whether it is the typical or the modified BE.

B. Training Sequence Allocation for DLS/MDLS Techniques

By taking a look at (25), it can be noted the MSE depends on the trace of the interfering covariance matrices of the users using the same sequences.

Theorem 1. *For typical LS, employing any training sequence allocation algorithm does not reduce the MSE performance.*

Proof. It can be proved from (25) that MSE performance depends on the traces of all involved covariance matrices and it does not depend on the signal space spanned by each correlation matrices. Therefore any set of correlation matrices has the same trace will result in the same MSE. □

However, this theorem does not apply for DCT based LS estimation. Since the MSE metric of LS does not provide us an indication about the spatial separation among the subspaces and the spatial frequencies. Since the correlation matrices capture the information about subspaces, the intersection of these subspaces results in interference. Therefore, to measure the spatial separation between the lth and m^{th} users towards

$$\mathcal{E}(\mathcal{U}_B(\mathbf{s}_l), \mathcal{U}_M(\mathbf{s}_1)) = \sum_{\forall c \in \mathcal{U}_M(\mathbf{s}_l)} tr\left(\mathbf{R}_{lc} - \mathbf{R}_{lc}^{\frac{3}{2}}\left(\sum_{\forall m \in \mathcal{U}_M(\mathbf{s}_l)} \mathbf{R}_{lc}^{\frac{1}{2}}\mathbf{R}_{mc}\mathbf{R}_{lc}^{\frac{1}{2}} + \frac{\sigma_n^2}{\tau}\mathbf{R}_{lc} + \zeta\mathbf{I}_M\right)^{-1}\right) \qquad (33)$$

$$+ \sum_{\forall c \in \mathcal{U}_B(\mathbf{s}_l)} tr\left(\mathbf{R}_{lc} - \mathbf{R}_{lc}^2\left(\sum_{\forall m \in \mathcal{U}_B(\mathbf{s}_l)} \mathbf{R}_{mc} + \frac{\sigma_n^2}{\tau}\mathbf{I}_M\right)^{-1}\right).$$

the cth BS, we define the following metric:

$$\delta_{lm}^c = \frac{tr(\mathbf{R}_{lc}\mathbf{R}_{mc})}{tr(\mathbf{R}_{lc})tr(\mathbf{R}_{mc})} \qquad (34)$$

where $0 \leq \delta_{lm}^c \leq 1$. When δ_{lm}^c is close to 1, the users span highly overlapped subspaces, but when δ_{lm}^c is close to 0, the users span a highly separated subspaces. The spatial separation between the mth user in c^{th} and interference subspaces of all interfering users at the estimation process can be written as

$$\delta_l^c = \frac{tr(\mathbf{R}_{lc}\sum_{m=1,m\neq l}^C \mathbf{R}_{mc})}{tr(\mathbf{R}_{lc})tr(\sum_{m=1,m\neq l}^C \mathbf{R}_{mc})}. \qquad (35)$$

Define the set of users who utilizing the training sequence $\mathbf{s}_l$ by $\mathcal{U}(\mathbf{s}_l)$, the set of all users $\mathcal{U}$ and the set of all users who have not assigned a training sequence yet as $\mathcal{M}$. We define the following metric as

$$\delta(\mathcal{U}(\mathbf{s}_l)) = \sum_{l \in \mathcal{U}(\mathbf{s}_l)} \frac{tr(\mathbf{R}_{lc}\sum_{m\in\mathcal{U}(\mathbf{s}_l),m\neq l}\mathbf{R}_{mc})}{tr(\mathbf{R}_{lc})tr(\sum_{m\in\mathcal{U}(\mathbf{s}_l),m\neq c}\mathbf{R}_{mc})}.$$

The algorithm can be summarized in (**A6**).

A6. Greedy Training Sequence Allocation Algorithm for DLS/MDLS

- Start from any training sequence. Without any loss of generality, we start with $\mathbf{s}_1$.
- Determine the training sequence reuse factor $\mathcal{K}$.
- Define the set of users who should allocate the same training sequence as $\mathcal{U}(\mathbf{s}_l)$
- Pick random user $u \in \mathcal{M} = \mathcal{U} - \cup_{n=1}^{l-1}\mathcal{U}(\mathbf{s}_n)$
- Step 1: set $\mathcal{U}(\mathbf{s}_l) = u$
- Step 2: if $|\mathcal{U}(\mathbf{s}_l)| \leq \mathcal{K}$ for users $\mathcal{M}(1)$ to $\mathcal{M}(|\mathcal{M}|)$
  1) $u^* = \underset{u \in \mathcal{M}}{\text{minimize}} \; \delta(\mathcal{U}(\mathbf{s}_l) \cup \{u\})$
  2) $\mathcal{M} = \mathcal{M}/\{u^*\}$
  3) $\mathcal{U}(\mathbf{s}_l) = \mathcal{U}(\mathbf{s}_l) \cup \{u^*\}$ go to Step 1.

However, the performance of DLS and MDLS can be optimized using training sequences allocation algorithm. This can be explained by their dependency on the spatial content of the users who utilize the same training sequence. Unfortunately, as theorem 1 states, the MSE metric of LS is incapable of addressing such a separation. On the other hand, the MSE of BE metric offers such a quality so we can optimize the performance of DLS and MDLS by allocating the training sequences as in previous greedy algorithms.

VIII. NUMERICAL RESULTS

In the previous sections, several algorithms were developed to decontaminate the training sequences in multiantenna multicell interference channels. In this section, the performance of these algorithms are evaluated and compared. Monte Carlo simulations are performed to examine the efficiency of the

LS	Least Square Estimator	(23).
BE	Bayesian Estimator	(18).
DBE	DCT Bayesian Estimator	(**A1**).
MBE	Modified Bayesian Estimator	(29).
DLS	DCT Least Square Estimator	(**A2**).
MDBE	Modified DCT Bayesian Estimator	(**A3**).
MDLS	Modified DCT Least Square Estimator	(**A4**).
ABE-MBE	Adaptive BE-MBE	(32).
ADBE-MDBE	Adaptive DBE-MBE	...
PA	Training sequence Allocation	**A5-A6**.
WPA	Without training sequence allocation	...

TABLE I
THE ACRONYMS USED IN THE SIMULATION SECTION FOR THE ADOPTED ALGORITHM AND THEIR CORRESPONDING EQUATIONS.

proposed algorithms. To assess the system performance, we need to compare it with the state of the art techniques as typical Bayesian estimator and least square estimator. The considered algorithms and the corresponding equations are summarized in table I.

We consider a multi-cell network where the users are all distributed around the base stations. We denote the ratio of the power of the direct link α_{cc} over interfering link α_{lc} with $\beta_{lc} = \frac{\alpha_{cc}}{\alpha_{lc}}$. To test the validity of the proposed algorithms, we deal with estimation at interference limited scenario, we consider $\beta_{lc} = 1, \forall l, c$, we drop the index β for the ease of notation. We adopt the model of a cluster of synchronized and hexagonally shaped cells. We assume that the training sequence $P = 0dB$. The used metric for evaluating the system performance is normalized sum mean square error and can be expressed as:

$$\epsilon = 10\log_{10}\left(\frac{\sum_{c=1}^C \|\hat{\mathbf{h}}_{lc} - \mathbf{h}_{lc}\|^2}{\sum_{c=1}^c \|\mathbf{h}_{lc}\|^2}\right). \qquad (36)$$

In this section, we adopt the correlation model in II-A2 θ_{sk} denotes the lower limit of the angular spread for each user with respect to the kth BS. $\Delta\theta$ is the span of the angular spread and it is assumed to be the same for all users. $\Delta_o\theta$ determines the amount of the overlap between the angular spread of different users. The compression ratio η is selected optimally using full search to minimize $\eta = \arg\min_{\eta\in(0,1]}\epsilon$ and this is considered along the simulations section unless mentioned otherwise.

A. Accurate Second order statistics

Fig. (4) illustrates the comparison among the different estimation techniques with respect to the number of antennas. It can be noted that MSE performance reduces monotonically with respect to the number of antennas at BS. It can be
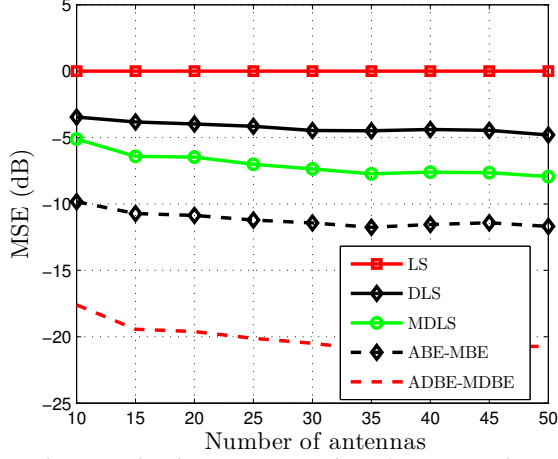
Fig. 4. *The normalized MSE ϵ vs. number of antennas, the considered scenario $C = 2$, $K = 2$, $\beta = 1$, $\Delta\theta = 20°$, $\theta_{s1} = \{10°, 25°\}$, $\theta_{s2} = \{20°, 35°\}$, $\Delta_o\theta = 5°$, and $P = 0$.*

compression ratio η. It can be noted that at high η, a large percentage of spatial frequencies are utilized. Therefore, the contamination is still contained in the influential frequencies of each estimate, which makes it hard to get an accurate estimate. On the other hand, when η is low, a small percentage of spatial frequencies are utilized which also removes a part of the useful signal to get an accurate channel estimate. Fig. (5) shows that for the DCT Based LS estimation (DLS, MDLS) the optimal η is lower than DCT Based Bayesian estimation (DBE, MDBE). This can be justified by the fact that typical BE estimation techniques, by their construction, have the capability of handling the interference at estimation in contrast to LS based estimation. This makes the DCT compression step for DLS and MDLS more valuable as it adds the capability of mitigating the contamination. While for DBE and MDBE, it enhances the capability of mitigating the interference, which makes it unnecessary to have lower values of η.

concluded that the LS has the worst performance in comparison with the other techniques. This can be explained by the fact that LS estimation just removes the impact of training sequence without introducing any processing to the aggregate of the received signals. DLS and MDLS are introduced to exploit the correlation information of the target estimate and incorporate the concept of DCT to get the important spatial frequency related to each estimate. It can be inferred that DLS and MDLS outperform the typical LS for all antennas scenario. In this figure, we also depict the comparison between the BE estimation techniques. The proposed techniques DBE and MDBE overcome the ABE-MBE. Intuitively, the adaptive MBE-BE performs better than BE and also this applies to DCT based technique, so the figures in this section display the comparison between the adaptive proposed schemes and the typical ones.
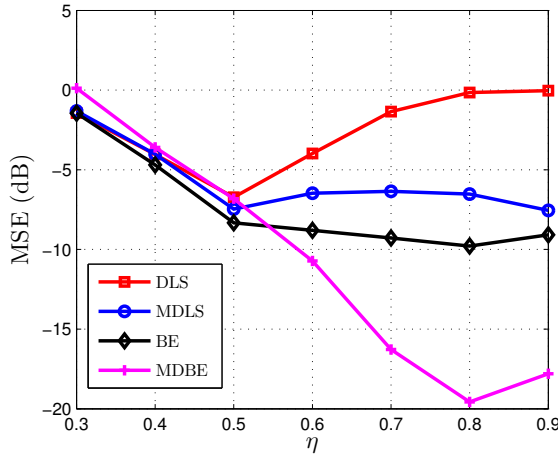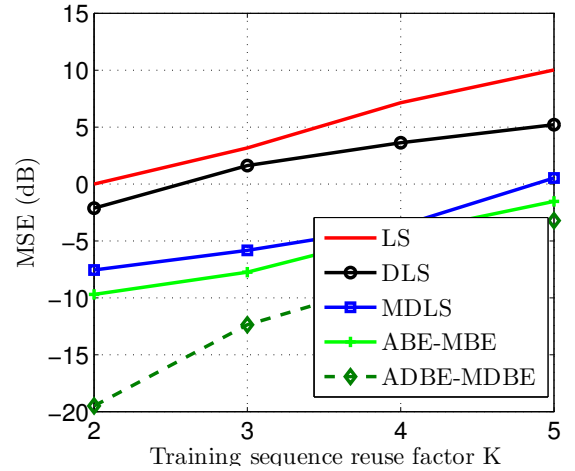

Fig. 5. *The normalized MSE ϵ vs. η, the considered scenario $C = 2$, $K = 2$, $\beta = 1$, $P = 0$, $\theta_{s1} = \{10°, 30°\}$, $\theta_{s2} = \{20°, 35°\}$ $\Delta\theta = 20$ and $\Delta_o\theta = 5°$.*

In Fig. (5), we study the performance of different proposed DCT based estimation techniques with respect to the

Fig. 6. *The normalized MSE vs training sequence reuse factors $M = 20, P = 0dB$, $\beta = 1$, $\theta_{s1} = \{10°, 25°, 40°, 55°, 70°\}$, $\theta_{s2} = \{20°, 35°, 50°, 65°, 80°\}$, $\theta_{s3} = \{0°, 15°, 30°, 45°, 60°\}$, $\theta_{s4} = \{35°, 50°, 65°, 70°, 85°\}$, $\theta_{s5} = \{85°, 70°, 55°, 40°, 25°\}$, $\Delta\theta = 20$, and $\Delta_o\theta = 5°$.*

Fig. (6) illustrates the MSE performance with respect to training sequence reuse factor. It is anticipated that increasing the training sequence reuse over the network increases the interference levels at the estimation, which makes it harder to obtain an accurate estimate of the required channel. This figure plots the comparison of different proposed estimation algorithms. It can be noted that there is a gap between ADBE-MDBE and ABE-MBE at low training sequence reuse factors, but this gap reduces with increasing the training sequence reuse factor. This can hamper the implementation of ADBE-MDBE at high training sequence reuse factor. However, the simulated scenarios consider interference limited case namely $\beta = 1$. MDLS performs closely to ABE-MBE for all training sequence reuse factors. Comparing MDLS with DLS, it can be seen that MDLS shows an enhanced performance over DLS, and the both proposed techniques outperform the typical LS for all training sequence reuse values.

Fig. (7) displays the comparison of different estimation techniques with the respect to the angular spread overlap $\Delta_o\theta$. The

considered scenario. Intuitively, higher overlap of the angular spread leads to higher training sequence contamination. It can be noticed that DLS and MDLS converge to LS in the scenario of complete angular spread overlap. While for BE techniques, it can be viewed that ADLS-MDLS has the same performance as ABE-MBE in case of complete angular spread overlap. For the scenario of distinct subspaces $\Delta_o\theta = 0°$, the typical BE and MBE and consequently their adaptive schemes perform better than ADBE-MDBE because there is no contamination at this scenario, and the DCT compression is meaningless. The
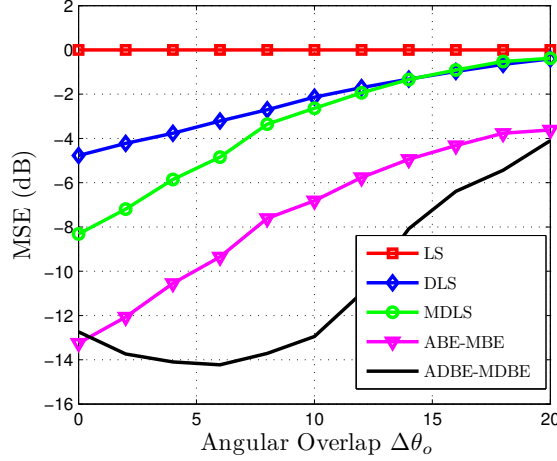

Fig. 7. *The normalized MSE vs angular overlap* $K = 2$, $C = 2$, $P = 0dB$, $\beta = 1$, $\theta_{s1} = \{10°, 30°\}$, $\theta_{s2} = \{70°, 85°\}$, $M = 10$.

efficiency of employing training sequence allocation algorithms is depicted in Fig. (8). We want to allocate 4 training sequences to 8 users in the two cells. The comparison of each estimation technique without and with training sequence allocation is plotted and it can be inferred that the system performance is enhanced if training sequence allocation algorithm is employed. This can be explained by the fact that selecting the which are assigned the same training sequence enhances the chances of being them naturally separable. Therefore, the contamination is reduced by employing these algorithms.

It is observed from the simulations that adapting the performance of different BE techniques as ABE-MBE and ADBE-MDBE achieves a better performance than employing just BE or MBE. Simulations have shown that the frequency of using MBE (or MDBE) versus BE (or DBE) in the adaptive algorithm is 75% versus 25% in the scenario of $C = 2$, $K = 2$, $\Delta_o\theta = 10°$ and $\Delta\theta = 25°$. This observation shows that the modified algorithm outperforms the original ones in majority of channel realizations.

B. The impact of inaccurate second order statistics

In the previous figures, we study the performance of the proposed techniques assuming accurate covariance acquisition at all BSs. Assuming inaccuracies and estimation errors in the covariance acquisition step may affect the performance of the suggested methods. The impact of these inaccuracies is depicted in Fig. (9), which plots the relation between the MSE and the uncertainty. It can be noted that typical
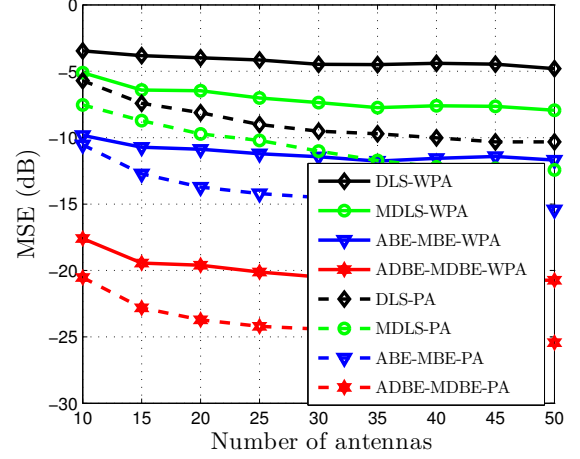

Fig. 8. *The normalized MSE vs number of antennas* $M = 20$, $P = 0dB$, $\beta = 1$, $\theta_{s1} = \{10°, 40°\}$, $\theta_{s2} = \{40°, 60°\}$, $\Delta\theta = 20$, *and* $\Delta_o\theta = 5°$.

LS is not affected by the uncertainty, this is intuitive since the estimation process does not depend on the covariance information. This fact also applies on the MSE assuming DLS estimation. Although the important spatial frequencies determination depends on the covariance information, the estimator design is independent from covariance information at BSs. Moreover, it can be concluded that the BE based techniques are more sensitive to covariance errors as they are functions of the covariance matrices of the involved users, the inaccuracies affect the contamination rejection at different BSs. These systems' MSE increase with respect to the amount of covariances inaccuracies. Finally, since MDLS is modified version of DLS and LS, which is based on increasing the covariance of the target channels, it is expected that the covariance inaccuracies degrade the estimation performance at different BSs.
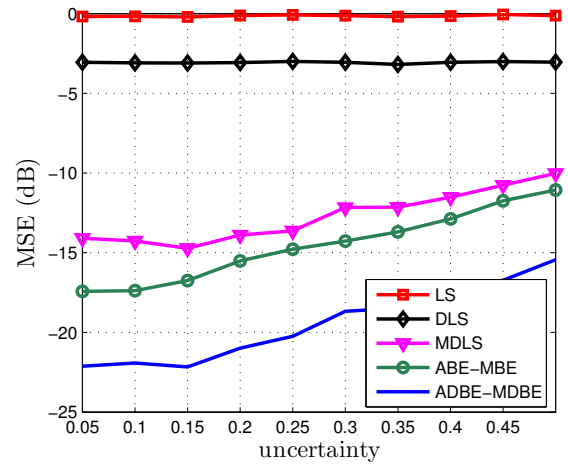

Fig. 9. *The normalized MSE vs number of antennas* $M = 20$, $P = 0dB$, $\beta = 1$, $\theta_{s1} = \{10°, 30°\}$, $\theta_{s2} = \{40°, 60°\}$, $\Delta\theta = 20$, *and* $\Delta_o\theta = 0°$.

The impact of inaccuracies considering the overlap in angular spread users' subspaces is depicted in Fig. (10). In compari-

son with Fig. (9), the performance of the proposed algorithms is studied. It can be noted that the proposed techniques are more sensitive to uncertainties which can be translated into higher MSE at all. As conclusion, to protect the system from these uncertainties, training sequence allocation should be adapted to take into consideration these uncertainties in their designs.
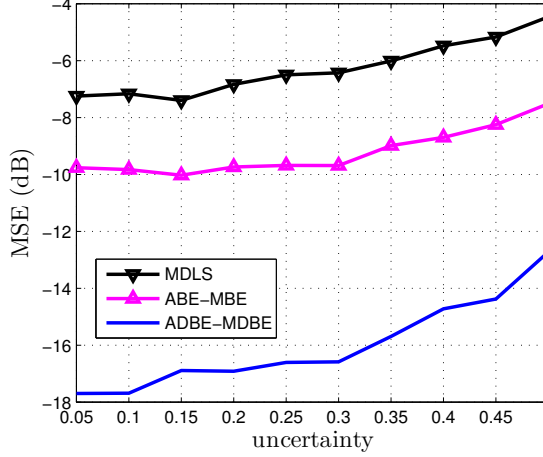

Fig. 10. *The normalized MSE vs number of antennas $M = 20, P = 0dB$, $\beta = 1$, $\theta_{s1} = \{10°, 40°\}$, $\theta_{s2} = \{40°, 60°\}$, $\Delta\theta = 20$, and $\Delta_o\theta = 5°$.*

IX. CONCLUSION

In this work, we studied the interference during the channel estimation and its impact on the system performance in multi-cell multiantenna networks. We investigated the performance of a Bayesian estimation and a least square estimation framework and formulated the lower bound and the upper bound of mean square error for such estimator. We proposed modified techniques to enhance the estimation accuracy by introducing the DCT, thereby transforming the problem into a different domain. This allowed the development of a new interference mitigation algorithm by compressing the spatial frequencies. It enabled enhanced estimation utilizing the DCT compressing capability to reduce the overlap in the interfering subspaces and boost the separation in the spatial domain. We incorporated this concept with training sequence allocation to assign different training sequences to reduce the interference in the overlapping area of the DCT subspaces. The performance of proposed algorithms was studied and compared to current state of the art techniques. From the simulation results, it can be concluded that the proposed algorithms provide considerable gains over the conventional Bayesian and least squares estimation techniques.

APPENDIX

Proof of Lemma 3: This can be proven by taking the expectation of $\mathbb{E}_\theta[\mathbf{h}_{lc}^d]$ at each spatial frequency k taking into the account the different distribution of the arrival angles at the BS. For uniform angular distribution on $[\theta_1, \theta_2]$

$$\mathbb{E}_\theta[\mathbf{h}_{lc}^d[k]] = \int_{\theta_1}^{\theta_2} \mathbf{h}_{lc}^d[k] dF_\theta = \frac{1}{\theta_2 - \theta_1} \int_{\theta_1}^{\theta_2} \mathbf{h}_{lc}^d[k] d\theta$$

Since the term $\omega_i = c\sin(\theta_i)$ exist in all k and make the integration term non integrable. For $k = 0$, the integration can be replaced by another integration,

$$\mathbb{E}_\omega[\mathbf{h}_{lc}^d[0]] = \frac{1}{c(\theta_2 - \theta_1)} \int_{\omega_1}^{\omega_2} \mathbf{h}_{lc}^d[0] \sqrt{1 - \left(\frac{\omega}{c}\right)^2} d\omega$$
$$\leq \frac{1}{c(\theta_2 - \theta_1)} \int_{\omega_1}^{\omega_2} \mathbf{h}_{lc}^d[0] d\omega \qquad (37)$$

For $k = 0$, If the angular spread angles close to zero, the previous term has a constant value, while if the angular spread is close to $\frac{\pi}{2}$ the previous integration is close to zero. For high frequencies, using the distributive property of integration, we can solve the related to integration by setting $x_1 = c\sin\theta - \frac{k\pi}{M}$ and $x_2 = c\sin\theta + \frac{k\pi}{M}$ as:

$$\mathbb{E}_\omega[\mathbf{h}_{lc}^d[k]] = \frac{1}{\theta_2 - \theta_1} \int_{c\sin\theta_1 - \frac{k\pi}{M}}^{c\sin\theta_2 - \frac{k\pi}{M}} \mathbf{h}_{lc}^d[k] \sqrt{1 - \left(\frac{x_1 + \frac{k\pi}{M}}{c}\right)^2} dx_1$$
$$+ \frac{1}{\theta_2 - \theta_1} \int_{c\sin\theta_1 + \frac{k\pi}{M}}^{c\sin\theta_2 + \frac{k\pi}{M}} \mathbf{h}_{lc}^d[k] \sqrt{1 - \left(\frac{x_2 - \frac{k\pi}{M}}{c}\right)^2} dx_2.$$

For angle of arrival close to zero, this integration will be close to zero. While for angle of arrival close to $\frac{\pi}{2}$, this term will result into considerable constant, which proves the lemma.

Proof of Lemma 4: If we assume θ_s is close to zero, finding the expectation for small angles close to zero at different spatial frequencies as

$$c(k) = \lim_{M\to\infty} \mathbb{E}_\omega[|\mathbf{a}[k]|] \qquad (38)$$

It can be found using (39)-(43) that the zero frequency converges to constant while the rest frequencies converge to zero. If θ_s is close to $90°$, all the frequencies will converge to zero, except for the highest spatial frequency.

REFERENCES

[1] E. Björnson, R. Zakhour, D. Gesbert and B. Ottersten,"Cooperative Multicell Pre-coding: Rate Region Characterization and Distributed Strategies with Instantaneous and Statistical CSI," *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4298-4310, August 2010.

[2] S. Chatzinotas, M. A. Imran, C. Tzaras, Capacity Limits in Cooperative Cellular Systems. Cooperative Wireless Communications, Auerbach Publications, Taylor and Francis Group, 2009.

[3] J. Zhang, and A. Jeffrey, "Adaptive spatial intercell interference cancellation in multicell wireless networks,"*IEEE Journal on Selected Areas in Communications*, vol. 28, no. 9, pp. 1457-1467, December 2010.

[4] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Transactions on Information Theory*, vol. 52, no. 11, pp. 5045-5060, November 2006.

[5] T. Yoo, N. Jindal, and A. Goldsmith, "Multi-antenna broadcast channels with limited feedback and user selection," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 7, pp. 1478-1491, September 2007.

[6] A. Ashikhmin and R. Gopalan, "Grassmannian packings for efficient quantization in MIMO broadcast systems," *IEEE International Symposium on Information Theory (ISIT)*, July 2007.

[7] N. Lee and W. Shin, "Adaptive Feedback Scheme on K-Cell MISO Interfering Broadcast Channel with Limited Feedback," *IEEE Transactions on Wireless Communications*, vol. 10, no. 2 pp. 401-406, February 2011.

[8] P. Zetterberg,"Experimental Investigation of TDD Reciprocity-Based Zero-Forcing Transmit Precoding," *EURASIP Journal on Advances in Signal Processing, Volume 2011, Article ID 137541*, 2011.

[9] F.A. Dietrich and W. Utschick, "Pilot-Assisted Channel Estimation Based on Second-Order Statistics," *IEEE Transactions on Signal Processing*, vol. 53, no. 3 , pp. 1178 - 1193, March 2005.

[10] E. Björnson and B. Ottersten, "A Framework for Training-based Estimation in Arbitrarily Correlated Rician MIMO Channels with Rician Disturbance," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1807-1820, March 2010.

[11] M. Biguesh, and A. B. Gershman, "Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals," *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 884-893, March 2005.

$$
c(k) = \begin{cases} \frac{0.5\lambda j}{2\pi d\Delta\theta}\left(\log\frac{1-e^{\frac{j2\pi d\theta_2}{\lambda}}}{1-e^{\frac{j2\pi d\theta_1}{\lambda}}} - \log\frac{e^{\frac{-j2\pi d\theta_2}{\lambda}}(-1+e^{\frac{-j2\pi d\theta_2}{\lambda}})}{e^{\frac{-j2\pi d\theta_1}{\lambda}}(-1+e^{\frac{-j2\pi d\theta_1}{\lambda}})} \right) + 1, & , M=odd, k=0, \\ \frac{0.5\lambda j}{2\pi d\Delta\theta}\left(\log\frac{1-e^{\frac{j2\pi d\theta_2}{\lambda}}}{1-e^{\frac{j2\pi d\theta_1}{\lambda}}} - \log\frac{e^{\frac{-j2\pi d\theta_2}{\lambda}}(-1+e^{\frac{-j2\pi d\theta_2}{\lambda}})}{e^{\frac{-j2\pi d\theta_1}{\lambda}}(-1+e^{\frac{-j2\pi d\theta_1}{\lambda}})} \right) & , M=even, k=0, \\ 0 & k \neq 0. \end{cases} \tag{39}
$$

$$
\mathbb{E}_\omega[\sqrt{M}|\mathbf{a}^{\text{DCT}}(0)|] = \begin{cases} \frac{\lambda}{2\pi d\Delta\theta}\left(\sum_{m=1}^{\frac{M-1}{2}} \frac{\sin m\theta_2 - \sin m\theta_1}{m} + \frac{2\pi d}{\lambda}(\theta_2-\theta_1) \right) & , M=odd, \quad k=0, \\ \frac{\lambda}{2\pi d\Delta\theta}\sum_{m=1}^{\frac{M}{2}}\frac{\sin m\theta_2 - \sin m\theta_1}{m} & , M=even, \quad k=0. \end{cases} \tag{40}
$$

$$
\sum_{m=0}^{\infty}\frac{\sin m\theta}{m} = \theta + 0.5j\left(\log\left(1-e^{j\theta}\right) - \log\left(e^{-j\theta}(-1+e^{j\theta})\right) \right) \tag{41}
$$

$$
\mathbf{a}^{\text{DCT}}(k) = \left(\left(\frac{\sin((\omega_i+\frac{\pi k}{M})\frac{M}{2})}{\sin((\omega_i+\frac{\pi k}{M})\frac{1}{2})}\right)^2 + \left(\frac{\sin((\omega_i-\frac{\pi k}{M})\frac{M}{2})}{\sin((\omega_i-\frac{\pi k}{M})\frac{1}{2})}\right)^2 + 2\frac{\sin((\omega_i+\frac{\pi k}{M})\frac{M}{2})}{\sin((\omega_i+\frac{\pi k}{M})\frac{1}{2})}\frac{\sin((\omega_i-\frac{\pi k}{M})\frac{M}{2})}{\sin((\omega_i-\frac{\pi k}{M})\frac{1}{2})}\cos\left(\frac{\pi(M+1)k}{M}\right) \right)^{\frac{1}{2}} \tag{42}
$$

$$
\mathbf{a}^{\text{DCT}}(k) = \begin{cases} \sqrt{2}\sqrt{\left(\frac{\sin((\omega_i+\frac{\pi k}{M})\frac{M}{2})}{\sin((\omega_i+\frac{\pi k}{M})\frac{1}{2})} - \frac{\sin((\omega_i-\frac{\pi k}{M})\frac{M}{2})}{\sin((\omega_i-\frac{\pi k}{M})\frac{1}{2})}\right)^2} \approx 0 & , M=odd, k\neq 0, \\ \sqrt{2}\sqrt{\left(\frac{\sin((\omega_i+\frac{\pi k}{M})\frac{M}{2})}{\sin((\omega_i+\frac{\pi k}{M})\frac{1}{2})} + \frac{\sin((\omega_i-\frac{\pi k}{M})\frac{M}{2})}{\sin((\omega_i-\frac{\pi k}{M})\frac{1}{2})}\right)^2} \approx 0 & , M=even, k\neq 0. \end{cases} \tag{43}
$$

[12] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions Wireless Communications*, vol. 9, no. 11, pp. 3590-3600, Nov. 2010.

[13] E. Björnson, M. Kountouris, M. Bengtsson, and B. Ottersten,"Receive Combining vs. Multistream Multiplexing in Downlink Systems with Multi-antenna Users,"*IEEE Transactions on Signal Processing*, vol. 61 , no. 13, pp. 3431 - 3446 ,July 2013.

[14] J. Jose, A. Ashikhmin, T. L. Marzetta and S. Vishwanath, "Pilot Contamination and Precoding in Multi-Cell TDD Systems," *IEEE Transaction Wireless Communications*, vol. 10, no. 8, pp. 2640 - 2651, Aug. 2011.

[15] K. Appaiah, A. Ashikhmin and T. L. Marzetta, "Pilot Contamination Reduction in Multi-user TDD Systems," *IEEE International conference on Communications (ICC)*, May 2010.

[16] H. Q. Ngo, T. L. Marzetta, and E. G. Larsson, "Analysis of the pilot contamination effect in very large multicell multiuser MIMO systems for physical channel models," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2011.

[17] B. Gopalakrishnan and N. Jindal, "An analysis of pilot contamination on multi-user MIMO cellular systems with many antennas," *IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2011.

[18] H. Q. Ngo and E. G. Larsson, "EVD-based channel estimation in multicell multiuser MIMO systems with very large antenna arrays," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2012.

[19] Y. Li, Y.-H. Nam, B. L. Ng and J. Zhang, "A Non-asymptotic Throughput for Massive MIMO Cellular Uplink with Pilot Reuse," *IEEE Global Conference on Communication (Globecom)*, December 2012.

[20] N. Krishnan, R. D. Yates, and N. B. Mandayam, "Cellular Systems with Many Antennas: Large System Analysis under Pilot Contamination," *Allerton Conference on Communications, Computing and Control*, October, 2012.

[21] H. H. Brunner, M. H. Castaneda, J. A. Nossek, "How Much Training is Needed for Interference Coordination in Cellular Networks," *ITG Workshop on Smart Antennas (WSA)*, March 2012.

[22] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A Coordinated Approach to Channel Estimation in Large-scale Multiple-antenna Systems," *IEEE Journal in selected Areas in Communications*, vol. 31 , no. 2, pp. 264 - 273, February 2013.

[23] M. Alodeh, S. Chatzinotas and B. Ottersten, "Joint Channel Estimation and Pilot Allocation in Underlay Cognitive MISO Networks," *invited paper to IEEE International Wireless Communications and Mobile Computing (IWCMC), Cyprus, August 2014.*

[24] A. Sayeed, " Deconstructing Multiantenna Fading channel,"*IEEE Transactions on Signal Processing*, vol. 50 , no. 10 , pp. 2563- 2579, October 2002.

[25] C. Martin, and B. Ottersten, "Asymptotic eigenvalue distributions and capacity for MIMO channels under correlated fading," *IEEE Transactions on Signal Processing*, vol. 3, no. 4, pp.1350 - 1359, July 2004.

[26] N. Ahmad, T. Natarajan and K. R. Rao, "Discrete cosine transform," *IEEE Transactions on Computers*, vol. 100, no. 1, pp. 90-93, 1974.

[27] H. Chen and B. Zeng, "New Transforms Tightly Bounded by DCT and KLT,"*IEEE Letters on Signal Processing*, vol. 19, no. 6, pp. 344 - 347, June 2012.

[28] Y.-H. Yehia and S.-G. Chen, "DCT-based channel estimation for OFDM systems," *IEEE International Conference on Communications (ICC)*, June 2004.

[29] P. Zetterberg and B. Ottersten, "The spectrum efficiency of a base station antenna array for spatially selective transmission," *IEEE Transactions on Vehicular Technology*, vol. 44, no. 3, pp. 57-69, August 1995.

[30] D. Katselis, C. R. Rojas, M. Bengtsson et. al, "Training Sequence Design for MIMO Channels: An Application-Oriented Approach," *EURASIP Journal on Wireless Communications and Networking*,2013 2013:245.

[31] A. Scherb and K. Kammeyer, "Bayesian channel estimation for doubly correlated MIMO systems," *IEEE Workshop on Smart Antennas (WSA)*, March 2007.

[32] S. M. Kay, Fundamentals of Statistical Signal Processing: Estimation Theory. Englewood Cliffs, NJ: Prentice Hall, 1993.

[33] A. N. Akansu and R. A. Haddad, *Multiresolution Signal Decomposition: Transforms, Subbands, and Wavelets.*Academic Press, Inc. 1992.

[34] F. Beaufays,"Transform-domain Adaptive Filters: An Analytical Approach," *IEEE Transactions on Signal Processing*, vol. 43, no. 2, pp. 422-431, February 1995.

[35] M. Alodeh, S. Chatzinotas and B. Ottersten, "Spatial DCT-Based Least Square Estimation n Multi-antenna Multi-cell Interference Channels," *IEEE International Conference on Communications (ICC), Sydney, Australia , June 2014.*